

```
library(papaja)

r_refs("library.bib")
```

We used R (Version 3.6.3; R Core Team 2020) and the R-package *papaja* (Version 0.1.0.9997; Aust and Barth 2020) for all our analyses.

Effects on perceived threat

To determine the effect of voice and race upon perceptions of threat, we ran a 2 (voice pitch: high or low) X 2 (race: Black or White) repeated measures ANOVA with perceived threat as the dependent measure (see Figure 1). There was a main effect of voice pitch upon perceived threat, $F(1, 506) = 62.225$, $p < .001$, $\eta^2_p = .11$. Race did not significantly predict perceived threat, $F(1, 506) = 0.170$, $p = .680$, $\eta^2_p = .000$, and the interaction between the variables was not significant, $F(1, 506) = 1.351$, $p = .246$, $\eta^2_p = .003$. We ran post-hoc tests with Bonferroni corrections to explore the main effect of voice pitch, which showed that low voices ($17.87 \pm .77$) were perceived as significantly more threatening compared to high voices ($12.34 \pm .61$), with a mean difference of 5.49(95% CI, 4.13 to 6.86).

Figure 1. Mean perceptions of threat as a function of voice pitch and perceived race. Error bars represent 95% confidence intervals. The perceptions of threat items were on 100-point scales.

Effects on perceived leadership ability

To determine the effect of voice and race upon perceived leadership ability, we ran a 2 (voice pitch: high or low) X 2 (race: Black or White) repeated measures ANOVA with the leadership composite score as the dependent measure. Both voice pitch, $F(1, 506) = 10.109$, $p = .002$, $\eta^2_p = .02$, and race, $F(1, 506) = 10.622$, $p = .001$, $\eta^2_p = .021$, significantly predicted leadership composite ratings (see Figure 2). Black voices ($61.72 \pm .73$) were rated higher on leadership traits compared to White voices ($59.12 \pm .72$), with a mean difference of 2.57 (95% CI, 1.03 to 4.16). On the other hand, lower voices ($61.72 \pm .74$) were rated higher on leadership qualities than higher voices ($59.12 \pm .72$), with a mean difference of 2.60(95% CI, .99 to 4.21). The interaction was not significant, $F(1, 506) = 0.079$, $p = .779$, $\eta^2_p = .000$. See Table A4 for mean differences in perceived leadership ability between White and Black voices based upon participants' demographic characteristics.

Figure 2. Mean perceptions of leadership traits as a function of voice pitch and perceived race. Error bars represent 95% confidence intervals. The perceptions of leadership items were on 100-point scales.

Trustworthiness and dominance predicting threat

The relationship between trustworthiness and dominance with perceived threat was examined by running a multiple linear regression with the averaged ratings across conditions. Overall, the predictors explained 19.7% of the variance in perceived threat, $F(2, 504) = 61.86$, $p < .001$, adjusted $R^2 = .194$. We found that trustworthiness was negatively related to threat, $b = -.28$, $t(504) = -6.72$, $p < .001$, while dominance was positively related to threat, $b = .44$, $t(504) = 10.46$, $p < .001$. We also ran the regression across all four conditions, confirming that the effect was present in each cell of our design (see Table A5).

Effects on trustworthiness and dominance

To determine the main effect of race upon perceived trustworthiness while controlling for the effects of voice pitch, we ran a 2 (voice pitch: high or low) X 2 (race: Black or White) repeated measures ANOVA with perceived trustworthiness as the dependent measure to examine whether race altered perceptions of

trustworthiness. There was a significant main effect of race, $F(1, 506) = 7.04, p = .008, \eta^2_p = .01$, upon perceived trustworthiness, where Black men ($61.93 \pm .70$) were perceived as more trustworthy than White men ($59.80 \pm .79$), while controlling for voice pitch and the interaction term. The mean difference in ratings was 2.13 (95% CI, 0.55 to 3.70).

We also tested whether voice pitch predicted perceived dominance independent of the effects of race by running a 2 (voice pitch: high or low) X 2 (race: Black or White) repeated measures ANOVA with perceived dominance as the dependent measure. Although the effect of voice pitch on perceived dominance was non-significant, $F(1, 506) = 3.49, p = .062, \eta^2_p = .007$, there was a significant main effect of race upon perceived dominance, $F(1, 506) = 68.25, p < .001, \eta^2_p = .12$. White men were perceived as more dominant ($46.15 \pm .94$) compared to Black men ($37.42 \pm .84$), with a mean difference of 8.73 (95% CI, 6.65 to 10.81).

Although some of the above tests failed assumption checks because they had outliers, we re-ran the tests without them and found the same effects. Therefore, we reported the results of the original tests with outliers included.

Exploratory analyses

Effects on individual leadership traits. We control for multiple hypothesis testing in all exploratory analyses by setting our criteria of significance at the .01-level. First, we ran a series of 2 (voice pitch: high or low) X 2 (race: Black or White) repeated measures ANOVAs with each of the leadership traits as dependent measures to break down the leadership composite effects. Participants were more likely to perceive Black recordings ($62.07 \pm .84$) as effective-communicators compared to White recordings ($58.64 \pm .82$), $F(1, 506) = 12.66, p < .001, \eta^2_p = .024$, with a mean difference of 3.43 (95% CI, 1.54 to 5.33). Also, participants were more likely to perceive low-pitched recordings ($58.02 \pm .82$) as problem-solvers compared to high-pitched recordings ($55.60 \pm .79$), $F(1, 506) = 7.63, p = .006, \eta^2_p = .02$, with an average difference of 2.42 (95% CI, 0.70 to 4.14).

There was a significant effect of voice pitch upon perceived confidence, $F(1, 506) = 20.94, p < .001, \eta^2_p = .04$, where low voices were perceived as significantly more than confident ($61.64 \pm .87$) than high voices ($56.74 \pm .83$), with an average difference of 4.90 points for the two groups (95% CI, 2.80 to 7.00). Also, race had a significant effect on perceived confidence, $F(1, 506) = 11.32, p = .001, \eta^2_p = .022$, where Black men were perceived as significantly more confident ($60.89 \pm .81$) than White men ($57.48 \pm .84$), with ratings differing by 3.41 points on average (95% CI, 1.42 to 5.39).

Main effect of race on perceived leadership ability

Given the unanticipated findings that Black men were rated as better leaders, we explore the data to examine three possible explanations for our outcomes: social demand effects, contrast effects, and the effects of stereotypes about Black men (i.e., more dominant and aggressive). To examine these possibilities for their validity in explaining our pattern of results, we will describe the evidence for and against each explanation based upon our exploratory analyses.

Social demand effects. First, social demand effects are always a concern when running within-subjects studies about race, because people are averse to being considered biased against Black people. When participants were presented 2 White names and 2 Black names (in a random order), it is entirely possible that they guessed that the study was focused on perceptions based upon race. Although we included a suspicion check and excluded participants based upon stringent criteria, the suspicion check may have biased participants to indicate that they were not suspicious. Specifically, they had to type in a text entry box if they were suspicious about the hypotheses, whereas they could simply select a multiple-choice option to indicate that they were not aware of the hypotheses. As a result, participants may have chosen the easier multiple-choice option on the suspicion check instead of choosing to type in their actual prediction of the purpose of the study.

To explore the plausibility of social demand effects as a potential explanation for our hypotheses, we tested a series of assumptions that we assumed would hold if participants were responding in a socially desirable way. If social demand effects were underlying our results, we would expect participants to rate the Black voices higher on leadership if they remembered the Black names better (i.e., performed better on the manipulation check). To test this assumption, we compared effects of the number of Black names that participants remembered (0, 1, or 2) upon the perceptions of leadership for the Black voices averaged across conditions by running a one-way ANOVA. This test suggested that performance on the manipulation check (i.e., memory for the Black names) did not significantly affect perceived leadership, $F(2, 504) = 1.00, p = .37, \eta^2_p = .004$.

The study also included interdependent relationship measures for exploratory analyses, where participants were asked how much they would like to engage in different types of interdependent relationships (i.e., work project team member, close friend, neighbor, employee) with the person in the recording using 100-point slider scale items. For these measures, we would also expect to see higher ratings of Black voices in relationships where Black people tend to be disadvantaged (e.g., employee or work project team member) because participants would try to avoid appearing biased. We ran a two-way (race by voice pitch) ANOVA with their preferences for having the recorded individuals as employees or work project team members. There was no significant effect of race on preferences for employees, $F(1, 506) = 3.36, p = .07, \eta^2_p = .007$, or work project team members, $F(1, 506) = 1.94, p = .16, \eta^2_p = .004$. This contradicts what we would expect for participants that are responding in any socially desirable way.

Other effects we would expect if participants were engaging in socially desirable responding are higher ratings for Black voices on trustworthiness and lower ratings on dominance and threat. Specifically, the prominent stereotype that Black men are criminals would prompt participants to rate them higher on trustworthiness if they did not want to appear biased. Along similar lines, perceptions of threat and dominance are a major stereotype that are applied to Black men (Quillian and Pager 2001), suggesting that participants should rate Black men lower on these traits to avoid appearing biased, if they correctly guessed our hypothesis. We found a significant effect of race upon perceived trustworthiness, $F(1, 506) = 7.04, p = .008, \eta^2_p = .014$, where Black men ($61.93 \pm .70$) were rated higher on perceived trustworthiness compared to White men ($59.80 \pm .80$), with a mean difference of 2.13 (95% CI, 0.55 to 3.70). Also, there were significant differences in ratings for perceived dominance between the races, $F(1, 506) = 68.25, p < .001, \eta^2_p = .119$, such that White men ($46.15 \pm .94$) were rated higher on perceived dominance compared to Black men ($37.42 \pm .84$), with a mean difference of 8.73 (95% CI, 6.65 to 10.81). These results provide support for socially desirable responding. However, contrary to what we would expect for participants that were trying to avoid responding in a biased manner, there was no significant effect of race on threat (see Figure 1 above). In sum, there is both evidence in favor of and against social demand effects in explaining the unexpected effect of race on perceived leadership.

Contrast effects. Another potential explanation for our unexpected results is contrast effects, which are based upon the shifting stereotypes model (Biernat, Manis, and Nelson 1991). This model posits that an individual will judge others on stereotype-relevant dimensions relative to other individuals within their social category. In the case of our study, the order of presentation of the name and voice stimuli may have affected the outcomes, since the stereotypical Black names were presented before the voices. This may have preempted them to expect a voice that sounded relatively uneducated. Previous research shows that the voice can convey SES and education levels (Kreiman and Sidtis 2011), so it is entirely possible that participants used this information in their assessments of the individuals in the recordings. Since the individuals that we recruited for our voice stimuli were generally well-educated (University of Pennsylvania graduate students and upper-level undergraduates) relative to the general population, the Black voices might have exceeded their low expectations, eliciting higher ratings. On the contrary, White men tend to have positive stereotypes attributed to them regarding their leadership ability (Rosette, Leonardelli, and Phillips 2008), so the baseline expectations for the White voices were relatively high, which may have also contributed to the effect of race upon perceived leadership ability.

If contrast effects can explain our results, we would expect to find a similar effect for Black men on the threat measure, where they would be rated significantly lower on perceived threat because their voice undermines the stereotype that Black men are extremely threatening. However, there is no main effect of race on perceived threat (see Figure 1). Since we created the leadership composite, it is entirely possible that it was

not a valid representation of the traits that are integral to a leader. However, our leadership composite is strongly correlated with the boss measure, $r(505) = .66$, $p < .001$, suggesting that it is a valid measure of leadership ability. Furthermore, we would expect participants to rate Black men higher on the boss measures if contrast effects contributed to our results. Contrary to this possibility, we find that participants did not exhibit any differential preferences based upon race, $F(1, 506) = 3.35$, $p < .068$, $\eta^2_p = .007$. Additionally, as the average ratings for perceived leadership increase, there should be a greater discrepancy between Black and White ratings, since we would expect contrast effects to be greater as the voices deviate more from what an individual expects based upon their stereotypes. We found this effect for almost all voices except for the voice that was rated highest on perceived leadership (see Figures A3 and A4).

Effects of stereotypes about dominance and aggressiveness. The final explanation that we explored in our data was the possibility that Black men were rated higher on leadership because, in the absence of threat, they may benefit from stereotypes that typically attribute dominance and aggressiveness to their social group (Devine and Elliot 1995). Dominance and aggressiveness were invaluable characteristics for leaders throughout our evolutionary history who needed to be successful during lethal inter-group conflicts and dangerous hunting sessions (Van Vugt, Hogan, and Kaiser 2008). Therefore, any personal characteristics that are perceived as more dominant or aggressive may increase perceived leadership ability, even though these traits may not accurately reflect leadership ability in the modern day (Klofstad and Anderson 2018; Li, Vugt, and Colarelli 2017). In this way, Black men might have been rated as better leaders because they were attributed dominance and aggressiveness to a greater degree than White men based upon stereotypes. If this explanation is valid, we would expect Black men to be rated higher on dominance, which, as demonstrated above, is not what we find in our data. Additionally, we would expect that participants would prefer Black men as a boss. Our analyses above do not support this assumption.

Aust, Frederik, and Marius Barth. 2020. *papaja: Create APA Manuscripts with R Markdown*. <https://github.com/crsh/papaja>.

Biernat, Monica, Melvin Manis, and Thomas E. Nelson. 1991. "Stereotypes and standards of judgment." *Journal of Personality and Social Psychology* 60 (4): 485–99. <https://doi.org/10.1037/0022-3514.60.4.485>.

Devine, Patricia G, and Andrew J. Elliot. 1995. "Are Racial Stereotypes Really Fading? The Princeton Trilogy Revisited." *Personality and Social Psychology Bulletin* 21 (11): 1139–50.

Klofstad, Casey A, and Rindy C Anderson. 2018. "Voice pitch predicts electability , but does not signal leadership ability." *Evolution and Human Behavior*. <https://doi.org/10.1016/j.evolhumbehav.2018.02.007>.

Kreiman, J, and D Sidtis. 2011. *Foundations of voice studies: An interdisciplinary approach to voice production and perception*.

Li, Norman P., Mark van Vugt, and Stephen M. Colarelli. 2017. "The Evolutionary Mismatch Hypothesis: Implications for Psychological Science." *Current Directions in Psychological Science*, 096372141773137. <https://doi.org/10.1177/0963721417731378>.

Quillian, Lincoln, and Devah Pager. 2001. "Black neighbors, higher crime? The role of racial stereotypes in evaluations of neighborhood crime." *American Journal of Sociology* 107 (3): 717–67. <https://doi.org/10.1086/338938>.

R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.

Rosette, Ashleigh Shelby, Geoffrey J. Leonardelli, and Katherine W. Phillips. 2008. "The White standard: Racial bias in leader categorization." *Journal of Applied Psychology* 93 (4): 758–77. <https://doi.org/10.1037/0021-9010.93.4.758>.

Van Vugt, Mark, Robert Hogan, and Robert B Kaiser. 2008. "Leadership, followership, and evolution." *American Psychologist* 63 (3): 182–96. <https://doi.org/10.1037/0003-066X.63.3.182>.